

Word Search Performance of Free-Text Electronic Patient Records by Surgical Diagnoses in a Veterinary Medical Teaching Hospital Clinical Database

Leah Estberg, DVM, MPVM, PhD[†]

James Case, DVM, PhD[†]

Richard Walters, PhD

Robert Cardiff, MD, PhD

Larry Gallupo, DVM

[†]California Veterinary Diagnostic Laboratory, University of California, Davis
Davis, California

Introduction. The University of California, Davis (UCD) Veterinary Medical Teaching Hospital (VMTH) clinical information system was developed to facilitate patient care and management. Much of the patient data can be entered as free-text in an electronic patient record (EPR) at a terminal keyboard, and individual patients and patient encounters are assigned unique identifiers. When the system is used to identify patients with searches on data other than these unique patient identifiers, issues of search accuracy must be considered such as search recall and precision. Search recall is defined as the percentage of all subjects of interest (cases) successfully retrieved ($[\text{cases retrieved during a search}] \div [\text{all relevant cases stored in the database}]$) and search precision as the percentage of all retrieved subjects that are cases ($[\text{cases retrieved during a search}] \div [\text{all subjects retrieved}]$). Methods are currently available to conduct word searches of the EPRs stored in the VMTH clinical database. The searcher can construct queries with multiple words combined with the boolean operators AND, OR and NOT, although search phrases are not supported. The objectives of the current project were to 1) identify limitations, if any, of search precision and recall for free-text surgical diagnoses included in electronic patient records maintained at the VMTH, 2) develop procedural or programmable recommendations for removing these limitations, and 3) provide guidelines for effective search strategies for users performing aggregate searches using the VMTH clinical information system.

Materials and Methods. Horses submitted to the UCD VMTH for a gastrointestinal disorder (colic) requiring surgical intervention from January 1, 1995 through December 31, 1995 were identified with the celiotomy procedure codes used for billing purposes and stored within the EPR.

Surgical reports for the case subjects were reviewed for major or primary surgical findings, and compared with the final Clinical Diagnoses listed in the EPR. Variations in expression of the main colic surgical findings reported as Clinical Diagnoses were tallied. Retrieval recall and precision were estimated for several gastrointestinal disorders in the study population. An equine surgeon at the VMTH (L. Gallupo) was provided with photographs of several gastrointestinal disorders discernable from reproduced photos (enterolithiasis, displacement, volvulus, adhesion) and asked to provide search terms he would use for record retrieval.

Results. Findings from a total of 227 hospital visits with 236 surgeries were included in the study. Retrieval performance was estimated for record searches with the terms enterolithiasis, displacement, torsion or volvulus, and adhesions. Search recall ranged from 33% to 98% and search precision ranged from 2% to 74%.

Conclusion. Successful patient aggregation by clinical diagnoses with word queries of free-text clinical information systems is largely influenced by the degree of variation in diagnoses description, level of detail provided, and familiarity with the local clinical vocabulary. Results from the current study demonstrated that successful identification of relevant cases, and *only* those cases, can vary greatly. The procedural recommendation that would likely have the greatest influence on minimizing these search limitations would be a movement towards uniform expression of gastrointestinal disorders. This would free searchers from having to imagine and anticipate all of the exact words and phrases that could be used in the relevant documents, and *only* in those documents.